

Writing in a Second Language with Machine Translation (WiLMa)

Margot Fonteyne^{1,2}, Maribel Montero Perez², Joke Daems¹, and Lieve Macken¹

¹ LT³, Language and Translation Technology Team

² MULTIPLES, Research Centre for Multilingual Practices and Language Learning in Society
Ghent University, Belgium

margot.fonteyne@ugent.be

Abstract

The WiLMa project aims to assess the effects of using machine translation (MT) tools on the writing processes of second language (L2) learners of varying proficiency. Particular attention is given to individual variation in learners' tool use.

1 Introduction

WiLMa (2021–2024) is a predoctoral research project funded by Ghent University's Special Research Fund (Grant No. BOF.DOC.2021.0001.01). The objectives for this project are:

- To compare the L2 MT-assisted writing process across proficiency levels with L2 writing processes not assisted by MT
- To map the individual variation in MT consultation behaviour (i.e., how L2 learners use MT during writing), investigate its correlation with learners' L2 proficiency level, and study its effects on the L2 writing product

The learners studied in this project are Dutch (L1) learners of Swedish (L2). In this project description, we report on the pilot study we carried out with these learners and present the next steps of this project.

Over the past decade, using MT has become a widespread practice among L2 learners, with writing tasks being one of the technology's most popular use cases (Jolley and Maimone, 2022). A number of publications has already investigated the effects of MT use on L2 writing using a product-oriented approach. These studies have shown that

writing products for which MT use was allowed differ from products for which it was not allowed.

However, our knowledge of the effects of MT use on the L2 writing process is still limited. By boosting learners' linguistic skills, MT may help learners to handle the competing demands on the different writing subprocesses better. These effects might also be larger for learners with lower proficiency levels (Révész, 2021). So far, two studies have investigated whether there are any differences to be found between learners' online writing behaviours (speed fluency, pausing, reading, and revising) in MT and non-MT conditions (Garcia and Pena, 2011; Raído and Torrón, 2020). However, it is difficult to draw conclusions from these studies, as they were based on a very small sample size and cover only a limited range of proficiency levels.

Moreover, despite consulting the tools being a major component of the MT-assisted writing process, few studies have investigated how learners use MT during writing. Cancino and Panes (2021) report for example that (untrained) learners look up 95 words per 100 words written. Fredholm (2015) notes that, on average, 44% of learners' texts is MT. However, these studies do not tell us whether learners' consultation behaviour when having access to MT is any different from when learners use more traditional writing tools, such as online bilingual dictionaries (OBDS).

Furthermore, research indicates that L2 learners' use of MT varies. This variation may be related to L2 proficiency (Fredholm, 2015) and likely also affects the learners' writing products (Cancino and Panes, 2021). By mapping the relationships between how learners with varying proficiency levels use MT and their writing products, we aim not only to find a (partial) explanation as to why MT-assisted writing products turn out to be

different from products for which MT was not allowed, but also to identify best practices for MT use by L2 learners of varying proficiency.

2 Pilot study

In this project's main experiment, we want to study the MT-assisted writing process across proficiency levels. To allow for this comparison, we need a reliable and valid instrument that quantifies learners' L2 proficiency level. We selected two tests that assess L2 learners' levels of Swedish: the Swedish Levels Test (Bokander, 2016) and a standardized placement test developed by Folkuniversitetet. Nine learners completed both tests. The internal consistency of the tests was high, as was the correlation between the learners' scores on the two tests. This correlation supports the criterion validity of the tests.

Moreover, we want to compare L2 learners' writing processes in two conditions: with access to an MT tool and with access to an OBD. To this end, the writing prompts the learners respond to should fulfill two criteria. First, they should elicit equivalent products and processes (i.e., be comparable). Second, they should be attainable to the least proficient learners, as well as challenging to the most proficient ones (i.e., be 'multilevel'). Therefore, we had 5 learners of varying proficiency respond to 4 prompts. In each prompt, we asked them to describe 3 images, choose the one that appealed to them the most, and explain why. This way, the texts contained both descriptive and argumentative elements, blending genres of varying difficulty.

Using the linguistic profiling tool Profiling-UD¹ and the keystroke analysis program Inputlog,² we analyzed the equivalence of the prompts by comparing product and process measures across the 4 tasks. The prompts elicited texts of similar length and complexity. The amount of time the participants spent on the tasks was comparable, as were their pausing and repair patterns. The participants consulted the tools that were allowed equally often and spent a similar amount of time consulting them. We also did not find any patterns in how difficult the learners perceived the different prompts.

The suitability of the prompts for use with multilevel learners was assessed by conducting similar analyses, but this time across levels instead of

prompts. The least proficient learners still managed to comfortably exceed the threshold of 100 words in the given time, which is needed to perform reliable automated analyses on the texts. The consultation measures show that even the most proficient learners relied heavily on the tools when responding to the prompts, indicating that the tasks were still challenging for them.

The data of this pilot study will be made available on OSF, licensed under CC BY-NC-SA.

3 Future work

In the future, we will collect data on the writing processes and products of multilevel learners, by having them respond to the piloted prompts in two conditions: with access to *DeepL* (MT) and with access to *Van Dale* (OBD). We will register their online behaviours with screen capture, keystroke logging, and eye-tracking, and their underlying cognitive processes with stimulated recall.

References

- Bokander, Lars. 2016. SweLT 1.0: konstruktion och pilottestning av ett nytt svenskt frekvensbaserat ordförrådtest. *Nordand*, 11(1):39–60.
- Cancino, Marco and Jaime Panes. 2021. The impact of Google Translate on L2 writing quality measures: Evidence from Chilean EFL high school learners. *System*, 98, article 102464.
- Fredholm, Kent. 2015. Online Translation Use in Spanish as a Foreign Language Essay Writing: Effects on Fluency, Complexity and Accuracy. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, 18:7–24.
- García, Ignacio and María Isabel Peña. 2011. Machine translation-assisted language learning: writing for beginners. *Computer Assisted Language Learning*, 24(5):471–487.
- Jolley, Jason R. and Luciane Maimone. 2022. Thirty Years of Machine Translation in Language Teaching and Learning: A Review of the Literature. *L2 Journal*, 14(1):22–44.
- Raído, Vanessa Enríquez and Marina Sánchez Torrón. 2020. Machine Translation, Language Learning and the 'Knowledge Economy'. In Michael Filimowicz and Veronika Tzankova (Eds.), *Reimagining Communication: Action* (pp.155–171). Routledge, New York, NY; Abingdon, Oxon.
- Révész, Andrea. 2021, May 14. *Investigating L2 writing processes: The roles of proficiency and stage of writing* [Conference Presentation]. LSLARF Annual Colloquium 2021, online. https://youtu.be/L2TcHbl_uaE?t=2007.

¹<http://www.italianlp.it/demo/profiling-ud/>

²<https://www.inputlog.net/>